



Analyse de mouvement : une revue

Soizic Laguitton, Christine Toumoulin

► To cite this version:

Soizic Laguitton, Christine Toumoulin. Analyse de mouvement : une revue. Innovation and Research in BioMedical engineering, 2009, 30 (2), pp.72-84. 10.1016/j.irbm.2008.12.007 . inserm-00678049

HAL Id: inserm-00678049

<https://www.hal.inserm.fr/inserm-00678049>

Submitted on 12 Mar 2012

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Analyse de mouvement. Une revue

Soizic Laguitton, Christine Toumoulin*

Inserm, U642, Rennes, F-35000, France

Université de Rennes 1, LTSI, Rennes, F-35000, France

Résumé

L'analyse des séquences temporelles d'images a fait l'objet de nombreux travaux depuis plus de vingt ans. Elle permet d'accéder à des informations importantes sur les objets contenus dans les scènes observées que ce soit à des fins de détection, d'estimation, de suivi ou de reconnaissance. Elle conduit aussi à de nombreuses applications dans des domaines très divers allant de la télévision à la météorologie, de la robotique à la surveillance ou encore au médical. Cette importance reconnue et la diversité des problématiques qu'elle recouvre ont motivé cet état de l'art qui se positionne dans un cadre générique plutôt qu'applicatif de façon à donner aux lecteurs une vue d'ensemble. Après avoir introduit les éléments clés à considérer (objets, observations, objectifs), trois grandes familles de méthodes sont analysées et discutées, les approches différentielles, les techniques bayésiennes et les méthodes de mise en correspondance. Un dernier paragraphe s'attache à faire le lien avec l'imagerie médicale en soulignant les spécificités qu'elle recouvre.

Abstract

Image sequence analysis has been a major topic for years. It allows accessing important information on the objects of interest for detection, estimation, tracking and recognition tasks. Many applications are concerned from video to weather forecasting, robotics to in-door monitoring up to medicine. The place that it has today and the diversity of the problems to address have motivated this state-of-the-art. A generic approach has been adopted in order to provide a full view on the recent advances reported in the literature at the methodological level. After introducing the key elements to be considered (objects, observations, objectives), three main classes of methods are analysed and discussed, the differential approaches, the Bayesian techniques and the matching methods. A last section is focused on the links with medical imaging with the aim to emphasize the special features to deal with.

Mots-clés : mouvement, détection, estimation, suivi, reconnaissance, modèle

Keywords: motion, detection, estimation, tracking recognition, modelling

* Corresponding author : LTSI, Campus de Beaulieu, Université de Rennes 1, 263 Avenue du Général Leclerc, CS 74205, 35042 Rennes Cedex, France

E-mail : christine.toumoulin@univ-rennes1.fr

I. Introduction

Prétendre faire un état de l'art exhaustif sur le mouvement en vision par ordinateur est aujourd'hui une gageure compte tenu de la multiplicité des travaux publiés. L'aborder par la seule fenêtre de l'imagerie médicale nous a semblé aussi trop peu informatif pour ceux qui peuvent demain s'intéresser à ce domaine. C'est donc d'un point de vue générique que nous abordons cette revue et, à ce titre, elle actualise les revues précédentes [1-2] et complète une étude récente consacrée au mouvement du corps humain [3] (plusieurs ouvrages de référence sont aussi disponibles [4-5]). Il convient dans un premier temps de donner quelques grilles de lecture à travers les différents regards portés sur l'analyse de mouvement décrits dans [6-8]. Une première grille conduit par exemple à mettre l'accent sur la nature des objets et leurs propriétés intrinsèques. Ce choix permet de souligner et de spécifier les connaissances a priori sur lesquelles certaines méthodes peuvent opérer. Une seconde voie consiste à privilégier la nature des données disponibles. Il est clair que le nombre de capteurs, l'accès direct ou pas à l'espace 3D, vont jouer un rôle important dans l'intérêt de telle ou telle approche. Une autre alternative est encore d'ordonner l'analyse de mouvement par grand problème générique comme la détection, la reconnaissance, etc.

Prenons le premier cas. La notion d'objets est ici considérée au sens large et concerne des objets artificiels, c'est-à-dire construits par l'homme, comme des objets présents dans la nature. Les frontières entre les différentes classes d'objets (rigides, articulés, déformables ou élastiques, fluides), ou par analogie les différents types de mouvements (articulés, élastiques, fluides), ne sont pas toujours aussi tranchées que nous pourrions le supposer. Suivant l'échelle de temps ou la résolution spatiale à laquelle il est étudié ou encore selon les contraintes auxquelles il est soumis, un objet rigide peut s'avérer déformable par exemple. Les *objets rigides* sont certainement les plus nombreux et a priori les plus simples à traiter. Leurs mouvements restent confinés à un ensemble de translations et de rotations dans l'espace 3D. Leurs apparences dans l'image du capteur, même en supposant celui-ci fixe, sont toutefois plus complexes à analyser car ils peuvent avoir des parties cachées changeant dans le temps (auto-occlusions), avoir des couleurs et des ombres portées différentes en fonction des conditions d'éclairage. Ils ont cependant l'avantage de posséder un mouvement cohérent, autrement dit tous les points de ces objets obéissent à la même loi de mouvement 3D. En projection cependant, outre le mouvement dans l'espace réel, le mouvement apparent (c'est-à-dire dans l'image) dépend de la distance (profondeur), de la position dans l'image et de la focale (effet de zoom). Les *objets articulés* sont définis comme des systèmes décomposables en parties distinctes (au moins deux) connectées entre elles par des articulations décrivant les mouvements relatifs possibles entre ces parties (six degrés de liberté au plus entre chacune). Le corps humain en est un très bon exemple par la multiplicité des applications qu'il engendre (du geste sportif à l'animation de personnages virtuels). En vision par ordinateur, avec ou sans marqueur et en général plusieurs caméras, le problème est de déterminer chaque partie dans l'image, d'estimer leur mouvement et d'identifier un modèle défini a priori du corps articulé [3]. Ce modèle doit être capable de représenter le corps dans toutes ses postures et rester suffisamment simple pour permettre une estimation précise et robuste de tous les paramètres du modèle. Plusieurs modèles géométriques des membres ont été proposés allant de primitives simples (cylindres) à des approximations polyédriques ou des superquadriques. Les stratégies d'identification procèdent soit en traitant chaque partie indépendamment des autres et puis les contraintes, soit en introduisant les contraintes dans le processus d'estimation global. Les problèmes de modélisation, d'initialisation et d'identification viennent ici s'ajouter aux difficultés liées aux occlusions. Les *objets déformables* ont des comportements élastiques (ils reviennent à leur forme initiale lorsque l'action exercée est supprimée), viscoélastiques (une partie de l'énergie est absorbée pendant la déformation), plastiques (déformation irréversible

après un choc sur un objet rigide). De nombreux modèles ont été proposés pour ces objets soit en implémentant les équations physiques de l'élasticité (linéaire ou non-linéaire), soit en construisant un assemblage d'éléments simples comme des masses et des ressorts. Notons que ces modèles déformables ont un impact important en segmentation d'image [9-10], en estimation de mouvement comme nous le verrons ensuite et en réalité virtuelle. L'une des applications les plus explorées dans ce cadre concerne l'analyse des expressions faciales s'intéressant aux mouvements et aux déformations de la tête, des lèvres, des yeux et des muscles faciaux [11]. La lecture sur les lèvres fait partie des sujets particulièrement traités car elle peut aider la perception acoustique dans des environnements très bruyés ou compléter le langage des signes pour les personnes handicapées [12]. Dans ce cadre, les méthodes doivent être robustes à l'échelle, la pose, l'éclairage et peuvent faire appel à la couleur, la texture, la forme et l'ombrage. Les domaines concernés par *les objets fluides* vont de la mécanique des fluides et l'aérodynamique aux sciences environnementales dont l'océanographie, l'étude des flux marins ou des dérives d'objets, la météorologie (reconstruction des vents à partir du déplacement des nuages). Les mouvements de ces objets fluides sont obtenus par la mesure des vitesses instantanées de particules fluides. Ils ont d'importantes variations spatiales et temporelles d'intensité avec le mouvement car il y a apparition et disparition de matière fluide et ne présentent pas de points remarquables stables. Comme dans les classes précédentes, des modèles des processus physiques sont exploités en particulier l'équation de continuité ou de conservation de masse [13]. L'hypothèse est d'une part que l'intensité dans l'image est supposée directement liée à une quantité passive transportée par le fluide et d'autre part que cette équation est valide en 2D pour le champ apparent (celui perçu dans l'image) de mouvement comme en 3D (mouvement réel). La caractérisation et l'interprétation du comportement d'un objet fluide est difficile dans la mesure où il n'a pas de support spatial au sens habituel: l'analyse de singularités dans les champs de mouvement ou d'évènements cinématiques sont alors essentiels pour comprendre les phénomènes observés et prédire leur évolution.

Dans le second cas, la nature et la quantité d'observations sont privilégiées. Compte tenu du large spectre des applications concernées par le mouvement, les principes physiques mis en oeuvre par les techniques d'observation, leurs modes et leurs conditions d'utilisation sont très divers. Il n'en reste pas moins que ce sont les capteurs optiques (les caméras dans le spectre visible) qui dominent en vision par ordinateur, de la télévision à la robotique, de l'imagerie sous-marine à l'endoscopie en médecine. Pour autant, des caractéristiques fondamentales et partagées se dégagent de l'ensemble des capteurs disponibles. Les séquences 2D sont les plus courantes où le capteur, en l'occurrence fixe, fournit une suite temporelle de projections 2D du monde 3D. Elle conduit donc à une perte d'information spatiale importante qui rend la plupart des tâches à réaliser plus ardues. Seuls les objets devant la caméra sont visibles. Les mouvements des objets perçus dans le champ de l'image sont des mouvements projetés et leur restitution (ou leur reconstruction) dans leur espace d'origine, lorsque l'application l'exige et en supposant la matrice de projection connue (via une calibration) va donc devoir faire appel à des informations supplémentaires. La première solution à ce problème est d'introduire des connaissances a priori sur les objets et leurs mouvements. La seconde est d'augmenter le nombre de capteurs: on parlera alors de séquences N-2D, le N représentant le nombre de capteurs. L'avantage d'un tel choix est de rendre plus robuste la plupart des algorithmes d'analyse en exploitant une redondance partielle mais aussi en réduisant le nombre de parties cachées et d'occlusions. L'inconvénient, outre le coût, est d'avoir à traiter des volumes de données plus conséquents. La vision stéréoscopique s'est développée dans cette logique. Elle a permis d'accéder sur la base de la géométrie épipolaire et par mise en correspondance de primitives à des informations 3D (bords ou surfaces d'objets dans des scènes d'intérieur pour la robotique par exemple). La précision en profondeur reste cependant réduite. L'utilisation de

systèmes multi-caméra, sans introduction de contraintes de faibles angles (comme c'est le cas pour l'analyse du mouvement humain), permet d'améliorer cette résolution mais au prix de rendre plus difficile la mise en correspondance (i.e. analyse sans marqueur optique). Le pas suivant est l'accès à des séquences volumiques indicées par le temps. La capacité d'imager des scènes dynamiques en volume fournit un avantage décisif puisqu'il s'affranchit des problèmes de reconstruction mentionnés ci-dessus. Cette situation favorable ne se retrouve malheureusement qu'en médecine et en biologie. Elle implique de grands volumes de données, et en conséquence de mémoire, ainsi que des temps de traitement loin du temps réel. Le troisième point de vue consiste à partir des objectifs poursuivis dans le cadre du problème posé. Ils peuvent concerner des tâches de bas niveau (détection ou estimation de mouvement) comme des problèmes de haut niveau (suivi, reconnaissance ou identification, ou encore reconstruction). Chacun de ces objectifs génériques, va s'instancier en fonction de l'application en exploitant la nature des objets, les modes d'observation, les connaissances ou modèles a priori éventuellement disponibles. La tâche de *détection* peut paraître à première vue simple puisqu'elle consiste à décider si, dans les images d'une scène, un ensemble de pixels (potentiellement réduit à l'unité) appartient ou pas à un objet en mouvement. La soustraction de deux images successives (e.g différence inter-image) peut offrir une solution. La détection d'intrusion dans des locaux surveillés par caméra fixe peut correspondre à ce type de situation. Le problème peut cependant s'avérer plus difficile si les images sont très bruitées auquel cas des approches plus élaborées de la théorie de la décision seront nécessaires. Il en est de même lorsque le capteur est mobile et qu'il faut détecter le mouvement d'un objet sans le confondre avec celui ou ceux du fond. Le pas suivant dans cet aspect détection consiste non seulement à le détecter mais aussi à étiqueter le type de changement comme par exemple "fond découvert par l'objet mobile", "fond recouvert par l'objet" ou "glissement de l'objet sur lui-même". L'*estimation* quant à elle vise calculer le mouvement apparent, ou mouvement projeté, en tout point de l'image ou, si une segmentation a été opérée auparavant, sur des primitives (contours ou régions par exemple). Réalisée de manière uniforme sur l'image, cette estimation va conduire à ce qu'il est convenu d'appeler des champs denses, ou flots optiques, de la vitesse instantanée de chaque pixel. Ces champs devront ensuite être analysés. Rapportée à des primitives supposées définir des éléments pertinents des objets présents, cette estimation évite un balayage complet, permet de filtrer des fausses alarmes et ajoute une information directement utile à la résolution du problème considéré (suivi, reconnaissance, etc.). Le bruit contenu dans les images va toutefois conduire à des estimations de mouvement localement biaisées et il faudra en général les lisser en utilisant des techniques de régularisation pour restituer des champs localement ou régionalement cohérents. Le mouvement peut aussi contribuer à la problématique de *segmentation* sous l'hypothèse que des objets distincts se déplacent de manière différente. L'objectif est alors de déterminer l'ensemble des vecteurs mouvements entre deux images successives ou sur un horizon temporel plus large et de les regrouper par zones homogènes. Lorsqu'un modèle paramétrique de mouvement (translationnel, affine, etc.) est imposé, ce modèle peut être identifié au moyen d'un sous-ensemble de vecteurs puis le regroupement de pixels être effectué sur la base de la compatibilité avec le modèle. Ce type de procédures peut être itéré pour améliorer l'estimation du modèle. Plusieurs difficultés intrinsèques à cette approche sont à souligner: les objets inhomogènes au sens du mouvement ou les objets immobiles ne peuvent pas être segmentés; déterminer l'intervalle temporel sur lequel travailler pour atteindre une segmentation consistante n'est pas évident. Nous voyons là, en tout état de cause, un questionnement possible dans l'articulation entre deux tâches différentes, segmentation statique et estimation de mouvement. Il porte sur l'ordre de réalisation de ces opérations mais aussi sur la fusion de leurs résultats. Le *suivi* et la *reconnaissance* d'objets font partie des finalités de l'analyse du mouvement et s'avèrent

potentiellement très imbriquées. La reconnaissance peut être envisagée en intra-image mais elle n'est pas exempte d'ambiguïté du fait des apparences multiples d'un objet selon le point de vue sous lequel il est regardé. Si cet objet est mobile et/ou si le capteur d'image est en mouvement, il est clair que l'accumulation d'information qui en résulte peut sensiblement réduire ces incertitudes. L'objet étant reconnu, des connaissances a priori peuvent être mises en oeuvre pour son suivi. De manière duale, un suivi procédant sans reconnaissance préalable peut aussi renseigner sur le type d'objet en fonction de la trajectoire qu'il suit. Quel que soit le cas de figure, l'objectif est de construire les trajectoires spatio-temporelles de ces objets. Il permet de prédire leur évolution à tout instant et donc d'anticiper sur des actions à entreprendre ou de compenser des périodes de croisement et d'occultation. Enfin, la *reconstruction* d'une forme (une structure 3D d'une scène dans un espace euclidien) a fait l'objet de nombreux travaux mettant l'accent sur la reconstruction à partir de l'ombrage ("shape from shading") ou du mouvement ("shape from motion"). Un tel objectif représente un intérêt majeur pour la manipulation d'objets, la navigation ou, plus généralement, pour toute tâche d'interaction en temps réel. La résolution de ce problème dépend tout d'abord des caractéristiques des objets, souvent géométriquement simples (assemblages de plans par exemple), des données d'observation (un ou plusieurs capteurs, fixes ou mobiles, calibrage connu ou pas) et de la pertinence des a priori retenus.

Cette décomposition entre objets, observations et objectifs permet d'appréhender toute problématique d'analyse du mouvement. Cependant, l'étude des grandes familles de méthodes d'analyse, à laquelle nous allons principalement consacrer cet article, reste cependant fondamentale. Nous avons retenu trois classes de méthodes : (i) les méthodes différentielles (section II) qui traitent les variations temporelles des intensités; (ii) les méthodes de suivi par filtrage bayésien (section III) exploitant des modèles probabilistes du mouvement et auxquelles nous consacrerons plus de place compte tenu des développements récents observés; (iii) les méthodes de mise en correspondance (subdivisées en mise en correspondance de primitives et mise en correspondance modèle/image) qui procèdent par recherche d'informations similaires dans des images successives (section 4). D'autres approches ne seront pas abordées malgré leur intérêt, comme les champs de Markov ou les techniques multirésolutions, car ils auraient nécessité des développements trop conséquents. Le dernier paragraphe de cet article s'attachera à discuter quelques uns des points clés de l'analyse de mouvement (hypothèses, modèles, limites, etc.) sur un plan général mais aussi relativement à l'imagerie médicale.

II. Les méthodes différentielles

Les méthodes différentielles sont issues en 1981 des travaux de Horn et Schunck [14], et de Lucas et Kanade [15]. Elles s'appuient sur l'hypothèse de conservation de la luminance au cours du temps considérée comme valide pour des mouvements de faible amplitude entre deux images successives [16-20]. La prise en compte de variations des conditions d'illumination, de réflexions spéculaires ou d'occlusions a cependant amené de nouvelles formulations plus générales en particulier celle de Papenberg et al. [21] :

$$\nabla(LI)^T \cdot w + \frac{\partial(LI)}{\partial t} = 0 \quad (1)$$

où $w = (u, v) = (\partial I / \partial x, \partial I / \partial y)$ est le vecteur vitesse recherché (le champ de vecteurs vitesse constituant le « flot optique ») et L désigne un opérateur chargé d'extraire de l'image l'information supposée invariante au cours du temps. Suivant le type d'image à traiter ou le type de mouvement, différents opérateurs peuvent être utilisés. Il peut s'agir simplement d'un filtrage visant à lisser le bruit et à obtenir ainsi une meilleure estimation des dérivées [22-23].

Cependant, l'utilisation de l'opérateur L a souvent comme objectif de maintenir l'équation (1.1) vraie en présence de variations d'illumination. Il est courant de modéliser simplement ces variations par une composante multiplicative et une composante additive $mI + a$ [21, 24-25]. Kim et Kak [24] utilisent directement l'opérateur $LI = mI + a$, a et m étant alors des paramètres à estimer en plus des vecteurs mouvements. Une autre solution pour faire face aux variations d'illumination consiste à utiliser des opérateurs différentiels ce qui permet d'éliminer au moins la composante additive. Les opérateurs gradient $L = \nabla$ et hessien $L = H$ supposent que les informations de direction sont conservées au cours du temps et sont donc adaptés aux mouvements de translation ou de divergence. Dans le cas de rotations importantes, il est possible d'utiliser la norme, la trace (laplacien) ou le déterminant du hessien, ou la norme du gradient [21]. Les dérivées d'ordre supérieur sont rarement utilisées car plus sensibles au bruit.

Les méthodes différentielles cherchent à estimer le mouvement (u, v) en chaque point, en résolvant le problème d'optimisation suivant :

$$(u, v) = \arg \min_{u, v} \left[\nabla(LI)^T \cdot w + \frac{\partial(LI)}{\partial t} \right]^2 \quad (2)$$

Cette formulation possède, cependant, trois inconvénients majeurs. Tout d'abord, la contrainte de flot optique n'est valable que pour des mouvements de faible amplitude. Ensuite, il existe presque toujours des points appelés "outliers" ou "aberrants" où cette hypothèse n'est pas vérifiée. C'est le cas, par exemple, en présence de mouvements multiples, de réflexions spéculaires, ou de changements d'illumination, lorsque ceux ci ne sont pas pris en compte dans la construction de l'opérateur L . Enfin, ce problème est en général sous-déterminé. C'est le problème de l'ouverture, largement évoqué dans la littérature, qui s'exprime ici : la variation d'intensité en un point est insuffisante pour déterminer le vecteur vitesse $w = (u, v)$. Des solutions pour chacun de ces problèmes ont été proposées. L'utilisation de schémas multirésolutions [23, 26-28] permettent non seulement d'étendre l'utilisation des méthodes différentielles à des mouvements d'amplitude importante, mais d'augmenter la robustesse au bruit, d'accélérer le processus et d'éviter les minima locaux. Pour éviter que la présence d'"outliers" ne perturbe le mouvement estimé, Black et Anandan [29-30] ont recours aux statistiques robustes qui s'attachent à développer des méthodes d'estimation fiables malgré la présence de points pour lesquels l'hypothèse de départ n'est pas vérifiée. Le problème d'optimisation est alors formulé ainsi :

$$(u, v) = \arg \min_{u, v} \rho \left(\nabla(LI)^T \cdot w + \frac{\partial(LI)}{\partial t} \right) \quad (3)$$

avec ρ , la fonction de l'erreur. L'influence qu'aura une certaine valeur de l'erreur est mesurée par une fonction appelée fonction d'influence ψ , proportionnelle à la dérivée de la fonction ρ . Dans le cas de la fonction "carrée", la fonction d'influence croît linéairement vers l'infini ce qui explique la vulnérabilité aux "outliers". Pour obtenir une fonction robuste, il est nécessaire de choisir une fonction ρ pour laquelle la fonction d'influence décroît pour des valeurs importantes de l'erreur, et tend ensuite vers zéro (Cf. [29] pour ces fonctions).

La question de l'indétermination est abordée par Olsen et Nielsen [31] par l'étude du flot normal aux gradients spatiaux de l'image. La plupart des travaux cependant [1, 22, 24, 32] s'appuient sur l'hypothèse d'un mouvement lisse en l'exprimant sous une forme globale, comme Horn et Schunck [14], ou locale suivant en cela Lucas et Kanade [15].

Dans le premier cas [14, 29, 33-36], une minimisation de fonction d'énergie globale est effectuée :

$$E_G = \int \int_{image} \left(\rho \left(\nabla(LI)^T \cdot w + \frac{\partial(LI)}{\partial t} \right) + \alpha V(\nabla u, \nabla v, \nabla I) \right) dx dy \quad (4)$$

où $V(\nabla u, \nabla v, \nabla I)$ est un terme de lissage pondéré par un scalaire α (Cf. Wieckert et Schnörr [34] pour une classification des termes de lissage et [22, 35, 37] pour quelques exemples). Les approches locales quant à elles supposent que le voisinage $\nu(s)$ du point s suive un même modèle paramétrique de mouvement. La formulation du problème est alors la suivante :

$$\hat{\theta}(s) = \arg \min_{\theta} \sum_{\nu(s)} \rho \left(\nabla(LI)^T \cdot w(\theta) + \frac{\partial(LI)}{\partial t} \right) \quad (5)$$

avec θ le vecteur des paramètres du modèle de mouvement, et $\hat{\theta}(s)$ le vecteur des paramètres estimés sur le voisinage $\nu(s)$ du point s . Pour éviter les erreurs aux frontières entre des objets de mouvements distincts, l'approche multimodale proposée dans Heitz et Bouthemy [38] ou une minimisation de la médiane des carrés [24, 27, 39] ont été proposées. Kim et Kak [24] ont décrit une variante de cette dernière méthode plus robuste au bruit. Au-delà du mouvement translationnel [22, 24, 28, 39, 40], les mouvements affine et quadratique sont traités dans [26] par un modèle polynômial à deux dimensions du second degré. Il apparaît, finalement, que le modèle affine est un bon compromis entre nombre de paramètres et réalisme du modèle [27].

Bruhn et al. en 2005 [22] ont confronté les approches locales et globales et proposé de les associer [23]. Les inconvénients de ces méthodes différentielles sont dus à plusieurs facteurs. Tout d'abord, elles se fondent sur l'hypothèse de conservation des fonctions de luminance, une hypothèse forte, et un compromis doit donc être fait entre l'exactitude de cette hypothèse et la complexité des fonctions de luminance construites. De plus, le problème est souvent mal posé et le lissage utilisé pour lever l'indétermination n'est pas adapté à la présence de mouvements multiples. Enfin, le calcul du mouvement s'appuie sur le calcul des dérivées, or ce calcul est souvent imprécis en raison de la présence fréquente de bruit ou d'échantillonnage spatio-temporel insuffisant. Il est rare que les méthodes proposées traitent tous ces problèmes simultanément en restant efficaces en termes de vitesse et de précision de calcul.

III. Suivi par filtrage bayésien

III.1 Principe

Les méthodes de suivi par filtrage bayésien modélisent l'état de la ou des cibles (objets suivis) par un processus aléatoire $X = (X_t)_{t \in N}$. Pour tout t , X_t est une variable aléatoire appelée vecteur d'état et contient classiquement la position de la cible ainsi que d'autres informations propres à chaque méthode. Sa dimension est notée N . Le filtrage bayésien permet d'estimer à chaque instant t , le vecteur d'état x_t en fonction d'observations y_t, y_{t-1}, \dots, y_0 partielles et bruitées de l'état, et des valeurs précédentes de l'état $x_{t-1}, x_{t-2}, \dots, x_0$. L'état est modélisé par un processus aléatoire markovien.

L'objectif est donc de déterminer à chaque instant, la densité de probabilité de x_t sachant y_t, y_{t-1}, \dots, y_0 . Elle sera ensuite utilisée pour calculer l'estimée \hat{x}_t , qui est en général le maximum de vraisemblance ou l'espérance. Le filtrage bayésien nécessite la construction de deux modèles stochastiques :

- Le modèle dynamique décrit l'évolution de la cible par une fonction f de l'état et du bruit blanc d'état V_t : $X_{t+1} = f(X_t, V_{t+1})$. Ce processus markovien peut, plus généralement être décrit par la densité de probabilité de transition $p(x_{t+1} | x_t)$.

- Le modèle de mesure décrit la relation entre l'observation Y_t et le vecteur d'état X_t , par une fonction g et un bruit de mesure W_t : $Y_t = g(X_t, W_t)$. Plus généralement, le modèle de mesure est décrit par la densité de probabilité $p(y_t | x_t)$.

Le calcul de la densité $p(x_{t+1} | y_{t+1}, y_t, \dots, y_0)$ est réalisé de manière récursive en deux étapes : (i) prédiction exploitant le modèle dynamique $p(x_{t+1} | x_t)$ et la densité à l'instant précédent $p(x_t | y_t, \dots, y_0)$ pour calculer la densité de probabilité de x_{t+1} sachant les observations précédentes y_t, \dots, y_0 ; (ii) correction donnant la densité de probabilité recherchée $p(x_{t+1} | y_{t+1}, \dots, y_0)$ à partir de la prédiction $p(x_{t+1} | y_t, \dots, y_0)$ et de la nouvelle observation y_{t+1} en utilisant le modèle de mesure $p(y_t | x_t)$. Dans le cas linéaire et additif gaussien, c'est à dire $X_{t+1} = AX_t + V_{t+1}$ et $Y_t = BX_t + W_t$, la probabilité $p(x_t | y_t, \dots, y_0)$ est gaussienne. Elle est déterminée par sa matrice de covariance et son espérance, calculées de manière exacte et récursive. Ce problème, résolu de manière optimale, est ce qu'on appelle le filtre de Kalman [41]. Dans le cas additif gaussien, avec des fonctions non linéaires, mais dérivables, c'est à dire $X_{t+1} = f(X_t) + V_{t+1}$ et $Y_t = g(X_t) + W_t$, il suffit de linéariser le système autour de la moyenne de l'estimée pour utiliser le filtre de Kalman. C'est le filtrage de Kalman étendu [42]. Une autre solution plus précise est le filtre de Kalman sans odeur [43]. Dans le cas où seulement une partie des composantes du vecteur d'état suivent un modèle conditionnellement linéaire et gaussien, l'évolution de ces composantes peut être calculée par filtrage de Kalman et un filtrage particulière appliqué aux autres composantes du vecteur d'état. Cette méthode est appelée "Rao-Blackwellisation". Le vecteur d'état peut aussi contenir des composantes à espace d'état fini. Dans ce cas, elles sont traitées de manière indépendante en utilisant des matrices de transitions [44].

Nous détaillons plus particulièrement le cas où les fonctions f et g sont fortement non linéaires où les méthodes de filtrage particulière prennent tout leur sens. L'idée clé est d'approcher la densité de probabilité $p(x_{t+1} | y_{t+1}, \dots, y_0)$ par des méthodes de Monte Carlo, dont le principe est le suivant. A condition qu'une loi p soit réalisable, l'espérance de la variable aléatoire $\varphi(X)$, $X \sim p$, peut être estimée à partir de M réalisations de la loi $\xi^i \stackrel{iid}{\sim} p(x)$ par :

$$I = \int_{\mathbb{R}^N} \varphi(x) p(x) dx \approx I^M = \frac{1}{M} \sum_{i=1}^M \varphi(\xi^i) \quad (6)$$

On appelle particules, les M réalisations de la loi (figure 1).

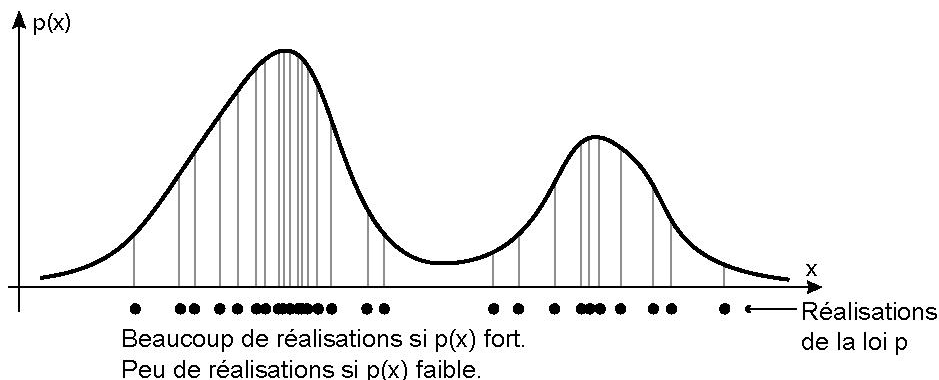


Figure 1 Illustration pour $N=1$ de l'approximation de Monte Carlo.

Dans le cas où la loi n'est pas ou est difficilement simulable directement il est possible de faire un échantillonnage d'importance. Une densité de probabilité simulable q est choisie, telle que $q(x) > 0 \Rightarrow p(x) > 0$: c'est la fonction d'importance. Les M réalisations sont alors tirées selon cette loi $\eta^i \sim q(x)$. L'approximation de l'intégrale est ensuite réalisée de la manière suivante :

$$I = \int_{R^N} \varphi(x) p(x) dx = \int_{R^N} \varphi(x) q(x) \frac{p(x)}{q(x)} dx \approx I^M = \frac{1}{M} \sum_{i=1}^M \varphi(\eta^i) \frac{p(\eta^i)}{q(\eta^i)} \quad (7)$$

Pour être efficace, la fonction d'importance choisie doit être proche de la loi $p(x)$ afin d'obtenir une bonne approximation à partir d'un nombre limité de particules (figure 2).

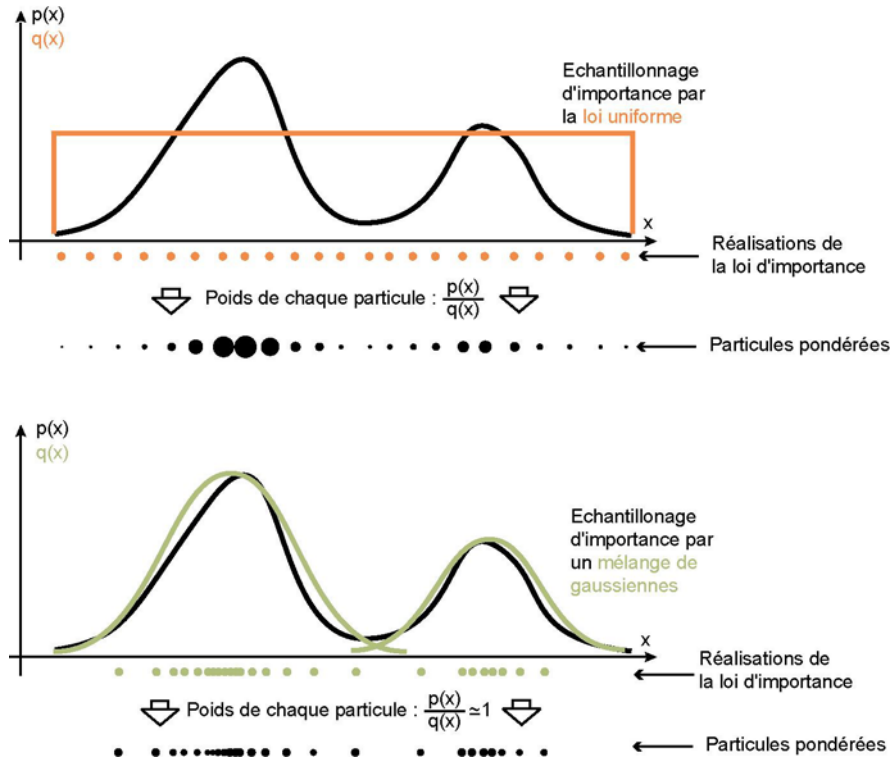


Figure 2 Illustration pour $N=1$ de l'échantillonnage d'importance.

Rappelons que le but est de calculer de manière récursive l'estimée \hat{x}_{t+1} , qui correspond à l'espérance de X_{t+1} conditionnellement à y_{t+1}, \dots, y_0 . L'utilisation de Bayes, puis de l'approximation de Monte Carlo utilisant une fonction d'importance $\pi(x_{t+1} | x_t, \dots, x_0, y_{t+1}, \dots, y_0)$ et les M particules $x_{t+1}^i \sim \pi(x_{t+1} | x_t, \dots, x_0, y_{t+1}, \dots, y_0)$, conduit à :

$$\begin{aligned}
\hat{x}_{t+1} &= \int_{R^N} x_{t+1} p(x_{t+1} | y_{t+1}, \dots, y_0) dx_{t+1} \\
&= \int_{R^N} x_{t+1} \frac{p(y_{t+1}, \dots, y_0 | x_{t+1}) p(x_{t+1})}{p(y_{t+1}, \dots, y_0)} \frac{\pi(x_{t+1} | x_t, \dots, x_0, y_{t+1}, \dots, y_0)}{\pi(x_{t+1} | x_t, \dots, x_0, y_{t+1}, \dots, y_0)} dx_{t+1} \\
&\approx \frac{1}{M} \sum_{i=1}^M \tilde{x}_{t+1}^i \frac{p(y_{t+1}, \dots, y_0 | x_{t+1}) p(\tilde{x}_{t+1}^i)}{p(y_{t+1}, \dots, y_0) \pi(\tilde{x}_{t+1}^i | \tilde{x}_t^i, \dots, \tilde{x}_0^i, y_{t+1}, \dots, y_0)} \\
&\approx \frac{1}{M} \sum_{i=1}^M \tilde{x}_{t+1}^i w_{t+1}^i
\end{aligned} \tag{8}$$

\tilde{w}_{t+1}^i est le poids attribué à la particule i au temps $t+1$. Ces poids peuvent être calculés de manière récursive en fonction du modèle d'observation, du modèle dynamique et de la fonction d'importance :

$$w_{t+1}^i = w_t^i \frac{p(y_{t+1} | \tilde{x}_{t+1}^i) p(\tilde{x}_{t+1}^i | \tilde{x}_t^i)}{\pi(\tilde{x}_{t+1}^i | \tilde{x}_t^i, \dots, \tilde{x}_0^i, y_{t+1}, \dots, y_0)} \tag{9}$$

Les deux problèmes principaux des méthodes de filtrage particulaire sont la dégénérescence des particules et la dégénérescence des poids. La dégénérescence des particules correspond à la migration de toutes les particules dans des zones où la probabilité $p(x_t | y_t, \dots, y_0)$ est très faible. La dégénérescence des poids correspond à la concentration des poids importants sur seulement quelques particules. Une solution pour détecter ce dernier phénomène est d'appliquer un seuil sur la variance des poids. Le problème peut être résolu en multipliant les particules de poids fort et en éliminant les particules de poids faible, de manière à retrouver M particules de poids $1/M$. Le ré-échantillonnage a été proposé par Gordon et al [45] en 1993 (SIR pour "Sampling Importance Resampling"). Cette méthode est utilisée dans pratiquement tous les travaux de filtrage récents appliqués au suivi de mouvement.

Les différentes méthodes de filtrage particulaire se distinguent principalement par la fonction d'importance choisie, c'est à dire la manière de faire évoluer les particules. La loi d'importance optimale correspond à $\pi(x_{t+1} | x_t, \dots, x_0, y_{t+1}, \dots, y_0) = p(x_{t+1} | x_t, y_{t+1})$. Elle permet de faire évoluer les particules en minimisant la variance des poids [46], ce qui limite les phénomènes de dégénérescence. Cette loi est, cependant, rarement disponible et simulable. L'approche la plus classique est d'utiliser la loi d'évolution (modèle dynamique) comme loi d'importance $\pi(x_{t+1} | x_t, \dots, x_0, y_{t+1}, \dots, y_0) = p(x_{t+1} | x_t)$. Associée à l'étape de ré-échantillonnage évoquée plus haut, cette méthode est appelée filtrage "bootstrap" [45]. L'inconvénient majeur est que les particules évoluent sans prendre en compte l'observation courante y_{t+1} . La méthode est donc très sujette au phénomène de dégénérescence et nécessite l'utilisation d'un nombre important de particules pour rester performante. Pour obtenir une fonction d'importance intégrant la dernière observation, une méthode répandue est le filtrage par particules auxiliaires proposé par Pitt et Shepard en 1999 [47]. Son principe est d'approcher, à chaque itération, la fonction d'importance optimale en simulant deux séries de particules.

III.2. Application au suivi de mouvement

Une partie des travaux les plus récents utilisent des modèles linéaires ou quasiment linéaires et donc le filtrage de Kalman simple ou sa version étendue [48-49]. Cependant, une place importante est donnée aux travaux utilisant des modèles plus complexes et non linéaires [44, 50-53]. Le filtrage particulaire a été appliqué pour la première fois à l'extraction de mouvement par Isard et Black en 1996, avec leur algorithme "condensation" [54]. Cette

méthode est en fait similaire à celle proposée plus tôt par Gordon et al [45], mais elle est appliquée au suivi de contours, et ne réalise pas d'étape de ré-échantillonnage des particules. Ce filtrage utilise la loi d'évolution comme fonction d'importance. Pour faire muter les particules en utilisant une information plus riche que la simple loi d'évolution, ces mêmes auteurs proposent en 1998 un algorithme nommé "ICondensation" [55]. Ils construisent une fonction d'importance à partir des informations fournies par un autre processus de suivi. A chaque itération, un tirage aléatoire détermine si la fonction d'importance utilisée est la loi d'évolution ou la nouvelle fonction. L'inconvénient est que la construction de ce deuxième processus n'est pas possible pour toutes les applications. De plus, le problème de la construction de la loi à partir des informations fournies par le second processus reste ouvert. Par la suite, beaucoup d'auteurs ont proposé des solutions pour obtenir une fonction d'importance proche de la loi optimale. Une solution efficace, proposée par Rui et Chen [56], est d'utiliser un pas du filtrage de Kalman sans odeur sur chaque particule, introduisant ainsi la dernière observation. La fonction d'importance est alors le mélange des gaussiennes obtenues. Le seul inconvénient de cette méthode est que la fonction d'importance doit être gaussienne, alors que la fonction d'importance optimale ne l'est pas en général. La méthode de filtrage par particules auxiliaires évoquée plus haut [47] est aussi une solution intéressante utilisée dans les travaux récents d'extraction du mouvement [57-58]. La recherche de méthodes permettant une exploration efficace de l'espace d'état et, donc la diminution du nombre de particules utilisées, reste un domaine actif [57-59]. Ces méthodes permettent, en effet, le suivi de mouvements dans le cas où le modèle dynamique est pauvre et la construction d'un vecteur d'état de dimension importante, et s'avèrent utiles dans le cas de cibles multiples [44, 58] ou l'utilisation de modèle articulé [50] par exemple.

Il nous reste à préciser maintenant les procédés de construction du vecteur d'état, de la mesure, des modèles dynamiques et de mesure.

Construction du vecteur d'état : La première fonction du vecteur d'état est de contenir l'information que l'on souhaite estimer. Il s'agit pour des applications au suivi, de la position 2D ou 3D de la ou des cibles. Il peut aussi inclure des informations spécifiques permettant de relier l'état courant à l'état précédent comme la vitesse [48, 60-61] ou des paramètres du modèle de mouvement utilisé [44, 62]. De plus, le vecteur d'état contient souvent des informations complémentaires spécifiques à chaque application, comme le nombre d'objets dans la scène dans le cas de cibles multiples [44] ou une variable pour la gestion explicite d'occlusions [58]. Dans le cas d'utilisation de modèles non linéaires, la construction du vecteur d'état doit réaliser un compromis entre la richesse des informations contenues et la complexité de la méthode. Les contraintes sont d'inclure dans le vecteur d'état des informations pertinentes et suffisantes pour un suivi efficace tout en conservant une dimension minimale pour limiter le nombre de particules nécessaires.

Construction du modèle dynamique : Le modèle dynamique doit inclure les connaissances a priori disponibles sur l'évolution du vecteur d'état. A notre connaissance, le bruit est toujours supposé gaussien, ce qui permet d'introduire de la souplesse dans le modèle. Il peut être avantageux d'utiliser une variance importante pour autoriser les mouvements rares et soudains (suivi de personnes par exemple [57, 61]). Les modèles les plus simples font l'hypothèse d'une faible variation du vecteur vitesse [60, 63]. Des modèles plus complexes peuvent être obtenus par apprentissage [62, 64].

Mesure-Détection : La mesure correspond dans la majorité des cas à une tâche de détection. La rapidité globale de la méthode dépend aussi de la vitesse de cette tâche. Une méthode très classique est la soustraction ou la modélisation du fond [53, 63] pour détecter les éléments en

mouvement. Cependant, cette méthode atteint ses limites en milieu naturel ou lorsque la caméra a un mouvement par exemple. De nombreuses solutions ont été employées : détection de bords [54], détection par la couleur [44], détection de points caractéristiques [48], détection de plans et de lignes [49], etc. Une approche moins classique consiste à utiliser une autre méthode de suivi (méthode de mise en correspondance) pour réaliser la mesure [19, 60].

Construction du modèle de mesure : Elle découle du choix de la mesure et du vecteur d'état. Le bruit du modèle de mesure est aussi généralement gaussien et permet de prendre en compte l'imprécision de la mesure réalisée mais aussi les fausses détections et les non détections dans le cas d'occlusions par exemple.

Les méthodes de suivi de mouvement par filtrage bayésien ont l'avantage d'utiliser conjointement les données images et un modèle de mouvement dont on peut régler la souplesse. La méthode peut être adaptée à tout type de modèles grâce aux méthodes de filtrage particulière. Leur inconvénient principal est la croissance exponentielle de leur complexité en fonction de la dimension du vecteur d'état, ce qui est une limitation dans le cas de modèles articulés ou de suivi multi-cible par exemple. Les méthodes actuelles tentent de résoudre ce problème en proposant de nouvelles techniques d'exploration de l'espace d'état, l'utilisation d'espaces d'état hybrides ("Rao-Blackwellisation" ou vecteur d'état contenant des variables aléatoires finies et infinies).

IV. Les méthodes de mise en correspondance

Les méthodes de mise en correspondance cherchent à trouver des informations similaires dans deux images espacées temporellement (ou spatialement dans le cas de la stéréovision). Les informations utilisées peuvent être : des primitives (points, lignes, cercles, ...), des modèles (déformables ou articulés par exemple) extraits de l'image, ou des informations statistiques sur les niveaux d'intensité des pixels (histogramme, entropie ...). Au delà de la simple mise en correspondance, l'objectif final de ces méthodes est le calcul de paramètres d'un modèle de mouvement paramétrique ou le suivi d'objet. Les méthodes utilisées sont généralement des méthodes d'optimisation ou d'ajustement des paramètres par rapport à une erreur mesurée (asservissement).

IV.1 Mise en correspondance d'un ensemble de primitives :

Nous nous intéresserons plus particulièrement au problème de mise en correspondance de points [65-69]. Le point est la forme la plus simple et à la base d'une construction de structures plus complexes comme des courbes et des surfaces. C'est la seule primitive invariante à tout type de transformation non rigide. Les marqueurs artificiels introduits dans le champ de vision sont généralement ponctuels comme des points dessinés physiquement sur le visage [65]. Ces points peuvent être aussi, selon la scène, des coins, des croisements. Par ailleurs, les méthodes proposées appliquées se généralisent le plus souvent aux autres types de primitives. La mise en correspondance d'un ensemble de points reste cependant un problème difficile. Des questions majeures se posent en effet : l'estimation conjointe correspondance-mouvement, l'élimination de points aberrants ("outliers") dans les données, et le traitement des déformations.

Estimation de la correspondance et du mouvement : certains travaux traitent les deux problèmes de manière séparée [65, 67, 70-71]. Les correspondances sont généralement déterminées en premier lieu (voire supposées connues a priori [67]). Etablir les correspondances de manière automatique est cependant une tâche non triviale [72]. Li et al.

[73] ont étendu l'approche de Kambhamettu et al [71] à des déformations plus importantes de maillages de surface. La cohérence de la correspondance de chaque point avec les points les plus proches du second maillage exploite des caractéristiques locales (vecteur normal, courbure, ...). Pour chaque point, le correspondant sélectionné est celui qui répond le mieux au critère établi. La réalisation de ce type méthode nécessite une forte connaissance a priori sur le mouvement. Cependant, dans la majorité des applications, les informations disponibles a priori sont insuffisantes. La stratégie généralement adoptée est alors de faire des hypothèses volontairement fortes, et d'utiliser une méthode d'estimation capable de gérer la présence de fausses correspondances (nous y reviendrons par la suite).

L'estimation conjointe correspondance-mouvement [48, 74-76] est plus élégante qu'une approche séquentielle. L'algorithme "ICP" pour "Iterative Closest Point", proposé en 1992 par Besl et al. [77] et largement appliqué pour le recalage, est la plus simple de ces méthodes. A chaque itération, chaque point transformé est assigné à son plus proche voisin, et cet appariement est utilisé pour raffiner les paramètres de mouvement. Cette opération est réalisée jusqu'à convergence. Cet algorithme est rapide, mais peut converger vers des minima locaux : il est donc sensible à l'initialisation. De plus, il s'avère peu robuste aux mauvaises ou fausses détections de points. Les travaux de Torr et al. [74] et Dellaert et al. [76] proposent des approches bayésiennes pour l'estimation conjointe des correspondances et des paramètres de transformation. Les observations sont les positions des points des deux ensembles et les images. Ces deux approches proposent de modéliser la vraisemblance (probabilité de l'observation sachant l'état) en supposant que l'erreur de position entre le point transformé et le point cible suit une loi normale de moyenne nulle. Elles diffèrent cependant dans la manière de réaliser l'estimation. Dellaert et al. utilisent un algorithme EM ("Expectation Maximization" [78]) quand Torr et al. mettent en oeuvre une approche multirésolution : la densité a posteriori (probabilité de l'état connaissant les observations) est approchée en réalisant un échantillonnage d'importance (cf. paragraphe III.1) par des particules pondérées par leur valeur de vraisemblance. Les particules sont propagées du niveau le plus grossier au niveau le plus fin par simulation de Monte Carlo par chaîne de Markov. Chui et Rangarajan [75] traitent l'estimation par optimisation d'une fonction de coût intégrant les correspondances et les paramètres de mouvement. La fonction pénalise la distance entre le point transformé et le point cible et favorise un mouvement lisse. Les correspondances sont répertoriées dans une matrice de correspondance qualifiée en logique floue. L'optimisation est réalisée par recuit déterministe où la température est fonction de l'entropie de la matrice de correspondance.

Gestion des données aberrantes : Les « outliers » sont inévitables du fait des limites des algorithmes de détection en présence de bruit et lorsque des scènes complexes (occlusions, variations d'illumination, etc.) sont analysées. Les statistiques robustes offrent des solutions pour traiter ces données aberrantes. Leur taux élevé en vision par ordinateur [79] a conduit à des développements spécifiques comme l'algorithme RANSAC ("Random Sample Consensus") proposé par Fischler et Bolles en 1981 [80]. Des travaux récents confirment son actualité [68-69] de même que ses évolutions depuis 2000 [70-74]. Son principe est le suivant : un sous-ensemble de données est construit de manière aléatoire; le modèle correspondant à ces données est estimé, et le nombre de points qui ne suivent pas ce modèle est évalué. Cette procédure est répétée un nombre fixé de fois. Le modèle donnant le moins de points aberrants (c'est-à-dire ne suivant pas le modèle) est sélectionné. Torr et Zisserman [70] ont amélioré cet algorithme en choisissant la solution qui maximise la vraisemblance au lieu de considérer simplement le nombre de points ne correspondant pas au modèle. Cette idée a été poursuivie par le même groupe pour des mouvements de forte amplitude en exploitant un schéma multirésolution [74]. L'inconvénient principal de ces méthodes est qu'elles ne sont pas adaptées aux transformations non rigides. Chui et Rangarajan [75] proposent en 2003 une

méthode générale basée sur les plaques minces ("Thin Plate Splines", [81]) permettant de traiter à la fois la présence de points aberrants et des mouvements déformables. Le problème est posé comme une minimisation d'une fonction d'énergie (cf. paragraphe précédent). Les points aberrants sont pris en compte en ajoutant une ligne et une colonne à la matrice de correspondance, chaque point étant soit assigné à un point de l'autre ensemble, soit attribué à la catégorie "outlier". Vogler et al. [65] ont proposé récemment plusieurs méthodes pour éliminer les correspondances aberrantes avant d'estimer les paramètres du mouvement. La première repose sur une étape de prédiction pour chaque point par filtrage de Kalman et l'élimination des correspondances trop éloignées de la prédiction. Cette méthode est simple mais nécessite de disposer d'un bon modèle d'évolution des points. La seconde utilise le même schéma mais en procédant dans l'espace des paramètres du mouvement.

Mouvements déformables : La mise en correspondance d'un ensemble de points dans le cas de mouvement déformable est un problème particulièrement complexe, et de plus en plus d'articles proposent des solutions [65, 67, 71, 73, 75]. Les mouvements déformables utilisés sont paramétriques, il peut s'agir de maillages [71, 73] avec une possible interpolation des points par splines [75] : les paramètres sont alors les positions des points de contrôle qui correspondent généralement aux points des ensembles à mettre en correspondance. L'inconvénient de ce type de modèles est qu'ils autorisent des déformations arbitraires et donc que pratiquement toutes les correspondances entre deux ensembles de points sont possibles. Des hypothèses supplémentaires sont donc nécessaires, typiquement des mouvements de petites amplitudes [71] et/ou des déformations faibles [71, 73, 75].

Un autre type de modèle paramétrique consiste à décomposer le corps déformable en une combinaison pondérée de formes de base [65-67]. Cette méthode est appelée factorisation de forme et tire son origine des travaux de Tomasi et Kanade en 1992 [82]. Ce type de modèle est par exemple appliqué aux visages dont la géométrie peut être vue comme la combinaison de plusieurs expressions de base. Les formes de base peuvent être obtenues par un apprentissage hors ligne, ou peuvent être estimées simultanément avec le mouvement [70].

IV.2. Mise en correspondance modèle/image

Dans ce cadre, un modèle, connu a priori ou segmenté sur la première image, est transformé à chaque instant en fonction des informations extraites de l'image. Un critère d'erreur entre le modèle et l'objet de l'image doit être construit. Ce critère est minimisé à chaque instant en prenant comme initialisation les paramètres du modèle à l'instant précédent. Dans le cas où le capteur (la caméra) est immobile, seuls les paramètres du modèle sont à déterminer à chaque instant [83-86]. Dans le cas où le capteur est mobile, il faut déterminer la position de l'objet dans l'image [87] ou la position du capteur [59]. L'application la plus représentative dans ce dernier cas est l'asservissement visuel [88-89] où l'objectif est de déplacer une caméra dans le but d'obtenir un point de vue spécifique de l'objet. Les points clés sont ici (i) la construction du modèle de l'objet, (ii) le choix des informations à extraire de l'image et du critère d'erreur, (iii) la méthode de minimisation retenue.

Modélisation de l'objet : Le cas le plus simple, l'*objet rigide* [59, 87], est décrit par un assemblage de primitives simples, comme des droites, points, cercles, cylindres et sphères, dont les relations géométriques dans l'espace 3D sont connues. Ce type de modèle peut être utilisé pour le suivi d'un objet dans l'image [87] ou l'estimation de la trajectoire du capteur lorsque l'objet est fixe [59]. Ces modèles conviennent particulièrement pour la modélisation de l'environnement humain (meubles dans une pièce, immeubles dans une rue) et trouvent

notamment des applications en réalité augmentée [87] nécessitant un traitement en temps réel. Les *objets articulés*, composés d'au moins deux éléments rigides reliés entre eux par des articulations contenant au plus six degrés de liberté, peuvent être des robots tels que ceux présents dans l'industrie [83] ou des objets divers, comme une boîte, un livre par exemple, souvent choisis pour tester les méthodes [83-84]. Le suivi du corps humain [50] ou d'une partie du corps humain, comme une main par exemple [59], est aussi souvent réalisé en utilisant ce type de modélisation. L'approche classique [59, 90] est de modéliser l'objet articulé par une chaîne cinématique. Les paramètres sont généralement stockés dans une structure de type arbre avec des informations sur la pose [50, 84], c'est-à-dire sur la position relative objet-caméra. Le problème est résolu comme un problème d'optimisation contraint par les relations d'articulations.

Pour les *objets déformables* en vision par ordinateur, le suivi de visages [85-86, 91-93] est un domaine particulièrement actif. Les modèles les plus employés sont : les modèles actifs d'apparence 2D ("Active Appearance Models") [94] et les "Morphable Models" 3D [95]. Ils se caractérisent par [86] :

- La forme, définie par la matrice f contenant l'ensemble des positions (2D ou 3D) des sommets d'un maillage
- L'apparence, définie par l'application $A(u)$ donnant l'intensité et la couleur (texture par exemple) de chaque point à la surface du maillage.
- Les paramètres de l'image q permettant de placer le modèle dans l'image. Il s'agit de paramètres de translation, rotation et homothéties pour les modèles 2D, et d'une matrice de projection pour les modèles 3D

La forme et l'apparence peuvent subir des transformations linéaires. Elles sont définies comme la somme d'une forme/apparence de base f_0/A_0 et d'une combinaison linéaire d'autres formes/apparences : f_i avec $i = 1, \dots, m$ et A_j $j = 1, \dots, l$:

$$\begin{aligned} f &= f_0 + \sum_{i=1}^m p_i f_i \\ A &= A_0 + \sum_{j=1}^l \lambda_j A_j \end{aligned} \quad (10)$$

Les formes et apparences f_0 , f_i , A_0 et A_j sont obtenues par apprentissage. Les formes et apparences de base, f_0 et A_0 , correspondent à la moyenne. Une analyse par composantes principales est ensuite réalisée, les formes f_i (respectivement apparences A_j) sont les vecteurs propres correspondant aux m (respectivement l) plus grandes valeurs propres. Les paramètres du modèle sont donc les coefficients $p = (p_1, \dots, p_m)$, $\lambda = (\lambda_1, \dots, \lambda_l)$ et les paramètres q , auxquels s'ajoutent, pour les modèles 3D, un modèle de réflectance comme le modèle de Phong [95]. Ces modèles sont utilisés pour le suivi de visages au cours du temps. Matthews et al. [96] proposent une approche temps réel avec modèle 2D. Les modèles 3D sont plus lents : Romdhani et al. [97] réalisent le suivi en 30 secondes par image. Ces derniers sont cependant plus robustes aux occlusions et convergent plus rapidement [86]. Pour les améliorer, Gross et al. [85] proposent d'augmenter la robustesse des modèles 2D en considérant les pixels occlus comme des "outliers" pris en compte par des méthodes statistiques robustes.

Exploitation de l'image et construction du critère d'erreur :

Lorsque le modèle représente uniquement la forme de l'objet, l'information de contour est utilisée dans la majorité des cas [83-84, 87, 98]. Elle présente l'avantage d'être invariante aux changements d'illumination. Pour augmenter la rapidité des méthodes, seuls les contours parallèles aux contours du modèle sont généralement recherchés [83-84]. Pour réaliser le

suivi, l'erreur à minimiser est alors la distance entre le modèle et le contour le plus proche dans l'image. Les modèles déformables de visage évoqués auparavant modélisent aussi l'apparence (c'est-à-dire les niveaux d'intensité et les couleurs). Le critère est alors la somme des différences au carré des valeurs des pixels de l'image avec les valeurs associées à chaque point du modèle [86]. Certaines approches couplent la caméra avec un autre type de capteur, par exemple un capteur de profondeur dans le cas de Bray et al. [59]. L'image est utilisée pour segmenter l'objet sur la base de sa couleur, la profondeur des points de l'objet est utilisée pour construire l'erreur entre le modèle et l'objet réel.

Minimisation du critère :

Le modèle et le critère d'erreur étant construits, l'objectif est de minimiser ce dernier par rapport aux paramètres du modèle. Les méthodes utilisées par les différents auteurs varient en fonction des contraintes spécifiques du problème (nombre de paramètres, espace non convexe...). Le domaine de l'optimisation étant vaste et complexe, nous prendrons uniquement comme exemple l'asservissement visuel virtuel spécifique au domaine de la vision. Dans le cas d'un objet rigide, le but est d'estimer sa pose à chaque instant par le biais d'une caméra virtuelle [84, 87, 99]. En notant P l'ensemble des positions 3D des différentes primitives composant le modèle de l'objet, r la position et l'orientation d'une caméra virtuelle en 3D, l'objectif est de minimiser l'erreur E entre l'ensemble des données observées s^* (généralement la position de l'ensemble des primitives dans l'image) et la position de ces mêmes données s obtenues par rétroprojection du modèle sur l'image connaissant la pose courante :

$$E = \sum_{i=1}^N (s_i(r) - s_i^*)^2 = \sum_{i=1}^N (retrop(r, P_i) - s_i^*)^2 \quad (11)$$

avec *retrop*, l'opération de rétroprojection. Comme en asservissement visuel 2D, la position et l'orientation de la caméra (ici virtuelle) sont modifiées par une boucle d'asservissement pour minimiser l'erreur E , l'ensemble des positions P étant fixe. A la convergence, la valeur de r correspond à la position et l'orientation réelle de la caméra. L'asservissement visuel virtuel offre tous les avantages des techniques d'asservissement visuel 2D actuelles : il permet le suivi d'objet en temps réel dans des scènes complexes, de manière précise, stable et robuste aux erreurs de mesure ou aux occlusions [87].

V. De la vision par ordinateur à l'imagerie médicale

La diversité des objets, des observations, des objectifs et des méthodes ne doit pas masquer leurs éléments communs. Dans la plupart des cas, les objets et leurs caractéristiques n'évoluent pas de manière soudaine. Les déformations restent progressives. De même, leurs mouvements sont plus souvent continus dans le temps qu'irréguliers ou, autrement dit, leurs trajectoires ne présentent pas de singularités. Cette continuité spatio-temporelle (en direction, en amplitude) peut être prise en défaut bien entendu (changement voire inversion de direction, changement de vitesse) et cette situation fait partie des sujets de recherche actuels. Les difficultés rencontrées avec les changements d'apparence se retrouvent dans toute problématique de reconnaissance d'objets en vision par ordinateur. La place des connaissances a priori est par ailleurs fondamentale. Celles-ci peuvent s'exprimer par des modèles d'objets (couleur, texture, forme), des modèles de comportement (mouvement, contraintes), soit construits par apprentissage statistique, soit fondés sur des lois physiques. Sans insister sur les problèmes de toute capture d'informations par l'image (le bruit et sa modélisation, les distorsions en particulier géométriques, la dépendance avec les conditions de prise de vue), les questions centrales sont d'une part la résolution temporelle et d'autre part

les degrés de liberté du capteur. Pour ce qui concerne le premier point, l'échantillonnage temporel détermine directement la capacité d'analyser des mouvements rapides de manière précise et valide, et les hypothèses qu'il est possible de faire sur le lissage ou l'amplitude du mouvement. Sur le second, une complexité supplémentaire s'ajoute si le ou les capteurs sont eux-mêmes mobiles (le fond de l'image n'est plus fixe) et, dans ce cas, disposer d'informations sur leurs positions au cours du temps devient essentiel.

Il est clair par ailleurs que de nombreux problèmes restent ouverts dès lors que des configurations difficiles s'accumulent: faible rapport signal à bruit, objets déformables, capteur(s) mobile(s), scènes complexes (incluant les objets multiples), occlusions, ruptures de mouvement, disparition-apparition de cibles, etc. C'est le cas du domaine du vivant, certainement un des champs d'application les plus exigeants pour l'analyse du mouvement. Il s'agit en grande majorité de mouvements complexes, irréductibles à des transformations rigides simples (comme la translation ou la rotation), où des mélanges de torsions globales et de déformations élastiques locales sont mis en jeu, en particulier pour le coeur en imagerie médicale que nous prendrons en exemple par la suite (en biologie d'autres processus augmentent ces difficultés [8]).

Si les techniques d'observation ont l'avantage d'être multiples et d'apporter des éclairages complémentaires sur les objets, elles mettent en jeu des processus physiques, des procédures d'acquisition et des méthodes de post-traitement très élaborés qu'il faut parfaitement maîtriser pour en tirer des indicateurs fonctionnels fiables. Leur point positif est qu'elles apportent des informations 3D directes que ce soit en ultrasons, en imagerie par résonance magnétique, en tomographie d'émission ou en tomodensitométrie X. De même, l'exploitation de produits de rehaussement ou de marqueurs extrinsèques permet d'accentuer la visibilité des organes cibles ou encore de tracer des échanges marquant un mécanisme particulier. L'inconvénient de ces marqueurs est malheureusement d'introduire une complexité supplémentaire dans l'analyse.

Les problèmes mentionnés auparavant de la détection à la reconstruction prennent tout leur sens en imagerie médicale. Les méthodes de flot optique par exemple ont eu leurs applications en angiographie X [100], en échocardiographie [101], en IRM (marqué) [102], et en scanner [103]. Il en est de même pour les méthodes de mise en correspondance [72, 104], et les modèles déformables [105-108]. Le suivi de marqueurs intrinsèques aux tissus par "speckle tracking" en échocardiographie occupe une place importante sur le plan clinique. Ces marqueurs sont en fait des diffuseurs réfléchissant une partie du signal initial. Le "speckle acoustique" provient donc de l'interférence des signaux rétrodiffusés par les inhomogénéités du milieu. Ces "speckle" (littéralement "tavelures" en français) vont se déplacer avec le mouvement du tissu, et leur suivi (par intercorrélation par exemple) permet donc une estimation du mouvement local. Les limitations de cette technique viennent de l'instabilité de ces diffuseurs et de leur déplacement en dehors du plan image observé. L'évolution vers le 3D de l'imagerie ultrasonore ouvre des perspectives nouvelles pour l'observation volumique du coeur, mais au prix d'un accroissement considérable des bases de données à stocker et à traiter.

L'exemple le plus proche des problématiques de vision par ordinateur en reconstruction concerne l'angiographie cardiaque par rayons X. Le caractère mal posé d'une reconstruction à partir de deux vues en angiographie X a motivé l'estimation du mouvement dans les séquences de projections et leur fusion en 3D. Une telle approche s'est révélée efficace dès lors qu'une première reconstruction était disponible [109]. Ce couplage mouvement-reconstruction apparaît essentiel aujourd'hui pour le Rotationnel X [110-111]. Les travaux du groupe de P.Grangéat [112] sur une reconstruction dynamique en présence de mouvement respiratoire sont certainement parmi les précurseurs de cette problématique qui trouvent aussi un intérêt majeur en tomographie par émission.

Le suivi d'objets hautement déformables dans des séquences volumiques comme en Scanner multi-détecteur ou en IRM présente un avantage certain par rapport à nombre d'applications évoquées précédemment, comme l'absence d'occultation. Cependant, le rapport signal à bruit, la présence d'artefacts, un échantillonnage temporel insuffisant viendront relativiser cet avantage. Par ailleurs, les scènes sont complexes par le nombre et la diversité des objets, leur proximité spatiale ou encore la similarité de leurs propriétés. De plus, les structures d'intérêt peuvent avoir des tailles faibles et très variables (de 1 à 5 mm de diamètre pour les artères coronaires par exemple), et sont soumises à des mouvements non stationnaires, présentant des ruptures importantes en direction et en vitesse.

Conclusion

Nous voyons après ce tour d'horizon partiel des méthodes d'analyse de mouvement à la fois la richesse des points de vue qu'elles offrent, la diversité des situations auxquelles elles s'adressent mais aussi leurs limites. Il souligne le rôle majeur joué par les modèles de mouvement et les approches bayésiennes. Le fait d'insister sur les difficultés ne signifie pas que les méthodes génériques décrites auparavant ne peuvent pas apporter des réponses pertinentes à un certain nombre de préoccupations exprimées dans le cadre de l'imagerie biomédicale. Il signifie cependant qu'il est particulièrement important d'exploiter les particularités des objets qui nous intéressent et les spécificités des modalités d'acquisition à notre disposition. La disponibilité de séquences volumiques suggère d'effectuer un suivi de mouvement directement en 3D, évitant ainsi les problèmes liés à l'apparition ou la disparition de points remarquables ou de zones d'intérêt dans les plans d'observation. Notons aussi que les performances que nous pouvons espérer sont très dépendantes des avancées techniques des modalités d'imagerie et que, ce qui est la vérité aujourd'hui, pour un objectif et une méthode de traitement d'image donnés, n'est pas forcément celle de demain. Autrement dit, il serait opportun de revisiter le comportement de nos algorithmes à chaque saut technologique afin de progresser dans l'extraction d'indicateurs fiables, robustes et reproductibles.

Références

1. Mitiche, A. and P. Bouthemy, Computation and analysis of image motion: a synopsis of current problems and methods. *Int. J. Comput. Vision*, 1996. 19(1): p. 29-55.
2. Orkisz, M. and P. Clarysse, Estimation du flot optique en présence de discontinuités : une revue. *Traitement du Signal*, 1996. 13(5): p. 489-513.
3. Moeslund, T.B., A. Hilton, and V. Kruger, A survey of advances in vision-based human motion capture and analysis. *Computer Vision and Image Understanding*, 2006. 104(2--3): p. 90-126.
4. Horaud, R., *Vision par ordinateur : outils fondamentaux*, Hermes, 1995.
5. Shapiro, L.G. and G.C. Stockman, *Computer Vision*, Prentice Hall, 2001.
6. Coatrieux, J.L., Shape and function from motion in medical imaging: Part I. *IEEE Engineering in Medicine and Biology Magazine*, 2005. 24(3): p. 29-31.
7. Coatrieux, J.L., Shape and function from motion in medical imaging: part 2. *IEEE Engineering in Medicine and Biology Magazine*, 2006. 25(1): p. 6-21.
8. Coatrieux, J.L., Shape and function from motion in biomedical imaging: part 3. *IEEE Engineering in Medicine and Biology Magazine*, 2007. 26(3): p. 81-83.
9. Kass, M., A. Witkin, and D. Terzopoulos. Snakes : Active contours models. in *Visual Computer* 1988, p. 321-331

10. Sethian, J.A., Level set methods and fast marching methods, Cambridge Univ.Press, 1999.
11. Kervrann, C. and F. Heitz, Modèles déformables statistiques pour l'analyse de structures 2D dans des séquences d'images. *Technique et Science Informatiques*, 1997. 16: p. 325--354.
12. Cladel, N., Optimisation multicritères de contours actifs par algorithmes génétiques : Application à la segmentation de la bouche. Thèse, Université de Rennes 1, 2005.
13. Corpetti, T., Estimation et analyse de champs denses de vitesses d'écoulements fluide, Thèse, Université de Rennes 1, 2002.
14. Horn, B. and B. Schunck, Determining Optical Flow. *Artificial Intelligence*, 1981. 17: p. 185--203.
15. Lucas, B.D. and T. Kanade. An iterative Image Registration Technique with an Application to Stereo Vision, *Proceedings DARA Image Understanding Workshop*. 1981, p. 121-130
16. Mitiche, A. and H. Sekkati, Optical Flow 3D Segmentation and Interpretation: A Variational Method with Active Curve Evolution and Level Sets. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2006. 28(11): p. 1818-1829.
17. Vazquez, C., A. Mitiche, and R. Laganier, Joint multiregion segmentation and parametric estimation of image motion by basis function representation and level set evolution. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2006. 28(5): p. 782-793.
18. Tagliasacchi, M., A genetic algorithm for optical flow estimation. *Image and Vision Computing*, 2007, 25(2), p 141-147
19. Arnaud, E., E. Mémin, and B. Cernuschi-Frias, Conditional Filters for Image Sequence-Based Tracking -- Application to Point Tracking. *IEEE Transactions on image processing*, 2005, 14(1), p 63-79
20. Ji, H. and C. Fermüller, A 3D shape constraint on video. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2006. 28(6): p. 1018-1023.
21. Papenberg, N., et al., Highly Accurate Optic Flow Computation with Theoretically Justified Warping. *International Journal of Computer Vision*, 2006. 67(2): p. 141--158.
22. Bruhn, A., J. Weickert, and C. Schnörr, Lucas/Kanade meets Horn/Schunck: combining local and global optic flow methods. *Int. J. Comput. Vision*, 2005. 61(3): p. 211-231.
23. Bruhn, A., et al., Variational optical flow computation in real time. *IEEE Transactions on Image Processing*, 2005. 14(5): p. 608-615.
24. Kim, Y.H. and A.C. Kak, Error Analysis of Robust Optical Flow Estimation by Least Median of Squares Methods for the Varying Illumination Model. *IEEE Trans. Pattern Anal. Mach. Intell.* , 2006. 28(9): p. 1418-1435.
25. Altunbasak, Y., R.M. Mersereau, and A. Patti, A fast parametric motion estimation algorithm with illumination and lens distortion correction. *IEEE Transactions on Image Processing*, 2003. 12(4): p. 395-408.
26. Odobez, J.M. and P. Bouthemy, Robust Multiresolution Estimation of Parametric Motion Models. *Journal of Visual Communication and Image Representation*, 1995. 6(4): p. 348-365.
27. Ong, E.P. and M. Spann, Robust Optical Flow Computation Based on Least-Median-of-Squares Regression. *Int. J. Comput. Vision*, 1999. 31(1): p. 51-82.
28. Mémin, E. and P. Pérez, Optical flow estimation and object-based segmentation with robust techniques. *IEEE Transactions on Image Processing*, 1998. 7(5): p. 703-719.

29. Black, M.J. and P. Anandan, The robust estimation of multiple motions: parametric and piecewise-smooth flow fields. *Computer Vision and Image Understanding*, 1996. 63(1): p. 75-104.
30. Black, M.J. and P. Anandan. A framework for the robust estimation of optical flow. in *Proc. Int Conf. Computer Vision 1993*, p.231-236
31. Olsen, O.F. and M. Nielsen, The Generic Structure of the Optic Flow Field. *Journal of Mathematical Imaging and Vision*, 2006. 24: p. 37--53.
32. Barron, J.L., et al., Performance of Optical Flow Techniques. *CVPR*, 1994. 92: p. 236-242.
33. Kim, Y.H., A. Martinez, and A.C. Kak, Robust motion estimation under varying illumination. *Image and Vision Computing*, 2005. 23(4): p. 365-375.
34. Weickert, J. and C. Schnorr, A Theoretical Framework for Convex Regularizers in PDE-Based Computation of Image Motion, *Int. J. Comput. Vision*, 2001, 45(3): p. 245-264.
35. Alvarez, L., J. Weickert, and J. Sanchez, Reliable Estimation of Dense Optical Flow Fields with Large Displacements. *Int. J. Comput. Vision*, 2000. 39(1): p. 41-56.
36. Aubert, G., R. Deriche, and P. Kornprobst, Computing Optical Flow Via Variational Techniques. *SIAM Journal on Applied Mathematics*, 1999, 60(1): p. 156-182.
37. Nagel, H.H. and W. Enkelmann. An investigation of smoothness constraints for the estimation of displacement vector fields from images sequences. in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1986, 8 (5), p 565-593
38. Heitz, F. and Bouthemy P, Multimodal estimation of discontinuous optical flow using Markov random fields. *IEEE Pattern Analysis Machine Intelligence*, 1993. 15(12): p. 1217-1232.
39. Bab-Hadiashar, A. and D. Suter, Robust Optic Flow Computation. *Int. J. Comput. Vision*, 1998. 29(1): p. 59-77.
40. Nesi, P., A.D. Bimbo, and D. Ben-Tzvi, A Robust Algorithm for Optical Flow Estimation. *Computer Vision and Image Understanding*, 1995. 62(1): p. 59-68.
41. Kalman, R.E. and R.S. Bucy, New Results in Linear Filtering and Prediction Theory. *Transactions of the ASME - Journal of Basic Engineering*, 1961. 83: p. 95-107.
42. Gelb, A., *Applied Optimal Estimation*, MIT Press, 1974.
43. Julier, S. and J. Uhlmann, A General Method for Approximating Nonlinear Transformations of Probability Distributions. 1996.
44. Czyz, J., B. Ristic, and B. Macq, A particle filter for joint detection and tracking of color objects. *Image and Vision Computing*, 2007, 25(8): p. 1271-1281.
45. Gordon, N.J., D.J. Salmond, and A.F.M. Smith, Novel approach to nonlinear/non-Gaussian Bayesian state estimation. *IEE Proceedings F Radar and Signal Processing*, 1993, 140(2): p. 107--113.
46. Doucet, A., On Sequential Monte Carlo Sampling Methods for Bayesian Filtering, *Statistics and Computing*, 1998, 10(3à, p 197-208.
47. Pitt, M.K. and N. Shephard, Filtering via Simulation: Auxiliary Particle Filters. *Journal of the American Statistical Association*, 1999, 94(446): p. 590-599.
48. Lippiello, V., B. Siciliano, and L. Villani, Adaptive extended Kalman filtering for visual motion estimation of 3D objects, *Control Engineering Practice*, 2007, 15(1): p. 123-134.
49. Dell'Acqua, A., A. Sarti, and S. Tubaro. 3D Motion from structures of points, lines and planes. in *Image and Vision Computing*, 2008, 26(4), p. 529-549.
50. Hua, G. and Y. Wu. A decentralized probabilistic approach to articulated body tracking. in *Computer Vision and Image Understanding*, 2008, 108(3), p. 272-283

51. Hansen, D.W. and R.I. Hammoud, An improved likelihood model for eye tracking. *Computer Vision and Image Understanding*, 2007, 106(2-3): p. 220--230.
52. Dowdall, J., I.T. Pavlidis, and P. Tsiamyrtzis, Coalitional tracking. *Computer Vision and Image Understanding*, 2007, 106(2-3): p. 205-219.
53. Lin, W.C. and Y. Liu, A lattice-based MRF model for dynamic near-regular texture tracking. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2007, 29(5), p 777-792
54. Isard, M. and A. Blake, Condensation : Conditional Density Propagation for Visual Tracking. *Int. J. Comput. Vision*, 1998, 29(1), p. 5-28.
55. Isard, M. and A. Blake, ICondensation: Unifying low-level and high-level tracking in a stochastic framework. *Lecture Notes in Computer Science*, 1998, 1406: p. 893-908.
56. Rui, Y. and Y. Chen, Better proposal distributions: object tracking using unscented particle filter. *Proceedings of the 2001 IEEE Computer Society Conference on computer Vision and Pattern Recognition*, 2001, 2, p. 786-793
57. McKenna, S.J. and H. Nait-Charif, Tracking human motion using auxiliary particle filters and iterated likelihood weighting. *Image and Vision Computing*, 2007, 25(6): p. 852-862.
58. Khan, Z., T. Balch, and F. Dellaert, MCMC Data Association and Sparse Factorization Updating for Real Time Multitarget Tracking with Merged and Multiple Measurements. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2006, 28(12), p. 1960-1972.
59. Bray, M., E. Koller-Meier, and L.V. Gool, Smart particle filtering for high-dimensional tracking. *Computer Vision and Image Understanding*, 2007, 106(1): p. 116-129.
60. Babu, R.V., P. Perez, and P. Bouthemy, Robust tracking with motion estimation and local Kernel-based color modeling. *Image and Vision Computing*, 2007, 25(8), p. 1205-1216.
61. Hu, W., et al., Principal Axis-Bases Correspondence between Multiple Cameras for People Tracking. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2006, 28(4), p 663-671
62. Cremers, D., Dynamical Statistical Shape Priors for Level Set-Based Tracking. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2006, 28(8), p 1262-1273
63. Munoz-Salinas, R., E. Aguirre, and M. Garcia-Silvente, People detection and tracking using stereo vision and color. *Image and Vision Computing*, 2007, 25(6): p. 995-1007.
64. Isard, M. and A. Blake. Contour tracking by stochastic propagation of conditional density. in *Proceedings European Conf. Computer Vision*, 1996, 1, p. 343-346
65. Vogler, C., et al., Outlier rejection in high-dimensional deformable models. *Image and Vision Computing*, 2007, 25(3), p. 274-284.
66. Zhu, G., Q. Zeng, and C. Wang, Efficient edge-based object tracking. *Pattern Recognition*, 2006, 39, p 2223-2226
67. Xiao, J., J. Chai, and T. Kanade, A Closed-Form Solution to Non-Rigid Shape and Motion Recovery. *International Journal of Computer Vision*, 2006, 67(2): p. 233-246.
68. Furukawa, Y., et al., Robust structure and motion from outlines of smooth curved surfaces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2006, 28(2): p. 302-315.
69. Wills, J., S. Agarwal, and S. Belongie, A Feature-based Approach for Dense Segmentation and Estimation of Large Disparity Motion. *International Journal of Computer Vision*, 2006, 68(2): p. 125-143.

70. Torr, P.H.S. and A. Zisserman, MLESAC: A New Robust Estimator with Application to Estimating Image Geometry. *Computer Vision and Image Understanding*, 2000, 78(1), p; 138-156
71. Kambhamettu, C., et al., 3D nonrigid motion analysis under small deformations. *Image and Vision Computing*, 2003, 21(3), p 229-245.
72. Simon, A., et al., Cardiac Motion Estimation in Multislice Computed Tomography Imaging Using a 4D Multiscale Surface-Volume Matching Process, *Proceedings Computers in Cardiology*, 2005, 32, p. 219-222
73. Li, M., C. Kambhamettu, and M. Stone, Nonrigid motion recovery for 3D surfaces. *Image and Vision Computing*, 2007, 25(3), p. 250-261.
74. Torr, P.H.S. and C. Davidson, IMPSAC: synthesis of importance sampling and random sample consensus. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2003, 25(3), p. 354-364.
75. Chui, H. and A. Rangarajan, A new point matching algorithm for non-rigid registration. *Computer Vision and Image Understanding*, 2003, 89, p. 114-141.
76. Dellaert, F., et al., EM, MCMC, and Chain Flipping for Structure from Motion with Unknown Correspondence. *Machine Learning*, 2003, 50(1-2), p. 45-71.
77. Besl, P.J. and N.D. McKay, A method for registration of 3-d shapes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1992, 14(2), p. 239-256.
78. Dempster, A., N. Laird, and D. Rubin, Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society*, 1977, 39(1), p. 1-38.
79. Meer, P., C.V. Stewart, and D.E. Tyler, Robust Computer Vision: An Interdisciplinary Challenge. *Computer Vision and Image Understanding*, 2000, 78(1), p. 1-7.
80. Fischler, M.A. and R.C. Bolles, Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography. *Comm. of the ACM*, 1981, 24, p. 381-395.
81. Duchon J, Splines minimizing rotation invariant semi-norms in Sobolev spaces, *Multivariate Approximation Theory*, Birkhauser, p 85-100, 1975
82. Tomasi, C. and T. Kanade, Shape and motion from image streams under orthography: A factorization method. *International Journal of Computer Vision*, 1992, 9(2), p. 137-154.
83. Comport, A.I., E. Marchand, and F. Chaumette, Kinematic sets for real-time robust articulated object tracking, *Image and Vision Computing*, 2007, 25(3), p. 374-391.
84. Drummond, T. and R. Cipolla, Real-time visual tracking of complex structures. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2002, 24(7), p. 932-946.
85. Gross, R., I. Matthews, and S. Baker, Active appearance models with occlusion. *Image and Vision Computing*, 2006, 24(6), p. 593-604.
86. Matthews, I., J. Xiao, and S. Baker, 2D vs. 3D Deformable Face Models: Representational Power, Construction, and Real-Time Fitting. *International Journal of Computer Vision*, 2007, 75(1), p. 93-113.
87. Comport, A.I., et al., Real-time markerless tracking for augmented reality: the virtual visual servoing framework. *Transactions on Visualization and Computer Graphics*, 2006, 12(4), p. 615-628.
88. Espiau, B., F. Chaumette, and P. Rives, A New Approach to Visual Servoing in Robotics, *IEEE Transactions on Robotics and Automation*, 1992, 8(3), p. 313-326.
89. Chaumette, F., Une nouvelle approche de la relation vision-commande en robotique, *Inria Report*, 1172, 1990.
90. Lowe, D.G., Fitting parameterized three-dimensional models to images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1991, 13(5), p. 441-450.

91. Hua, G. and Y. Wu, Sequential mean field variational analysis of structured deformable shapes. *Computer Vision and Image Understanding*, 2006, 101(2), p. 87-99.
92. Wang, J. and L. Yin, Static topographic modeling for facial expression recognition and analysis. *Computer Vision and Image Understanding*, 2007, 108(1-2), p. 19-34.
93. Tu, J., H. Tao, and T. Huang, Face as mouse through visual face tracking. *Computer Vision and Image Understanding*, 2007, 108(1-2), p. 35-40.
94. Cootes, T.F., G.J. Edwards, and C.J. Taylor, Active appearance models. *Transactions on Pattern Analysis and Machine Intelligence*, 2001, 23(6), p. 681-685.
95. Blanz, V. and T. Vetter. A morphable model for the synthesis of 3D faces. in *Annual Conference Series In Computer Graphics (SIGGRAPH)*, 1999, p. 187-194
96. Matthews, I. and S. Baker, Active Appearance Models revisited. *International Journal of Computer Vision*, 2004, 60(2), p. 135-164.
97. Romdhani, S. and T. Vetter. Efficient, robust and accurate fitting of a 3D morphable model. in *Proceedings of the International Conference on Computer Vision*, 2003, 1, p.59-62
98. Bazin, P.L. and J.M. Vezien, Integration of Geometric Elements, Euclidean Relations, and Motion Curves for Parametric Shape and Motion Estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2005, 27(12), p. 1960-1976.
99. Marchand, E. and F. Chaumette, Virtual Visual Servoing: a framework for real-time augmented reality, in *Proceedings EUROGRAPHICS'02 Conference*, 2002, p. 236-245.
100. Rong, J.H., J.L. Coatrieux, and R. Collorec, Motion estimation in digital subtraction angiography. *Proceedings Annual International Conference of the IEEE Engineering*, 1989, 2, p. 567-568.
101. Mailloux, G.E., et al., Computer Analysis of Heart Motion from Two-Dimensional Echocardiograms. *IEEE Transactions on Biomedical Engineering*, 1987, 34(5), p. 356-364.
102. Zerhouni, E.A., et al., Human heart: tagging with MR imaging--a method for noninvasive assessment of myocardial motion. *Radiology*, 1988, 169(1), p. 59--63.
103. Song, S.M. and R.M. Leahy, Computation of 3-D Velocity fields from 3-D Cine CT Images of a Human Heart, *IEEE Transactions on Medical Imaging*, 1991, 10(3), p. 295-306
104. Prince, J.L. and E.R. McVeigh, Motion estimation from tagged MR image sequences. *IEEE Transactions on Medical Imaging*, 1992, 11(2), p. 238-249.
105. Garreau, M., et al., Assessment of Left Ventricular Function in Cardiac MSCT Imaging by a 4D Hierarchical Surface-Volume Matching Process, *International Journal of Biomedical Imaging*, 2006, 1, p. 1-10
106. Clarysse, P., D. Friboulet, and I.E. Magnin, Tracking geometrical descriptors on 3-D deformable surfaces: application to the left-ventricular surface of the heart, *IEEE Transactions on Medical Imaging*, 1997, 16(4), p. 392-404.
107. Clarysse, P., D. Friboulet, and I.E. Magnin. Curvature assessment of the heart walls in 3D dynamic imaging, *Proceedings Computers in Cardiology*, 1993, p.607-610
108. Young, A.A. and L. Axel, Three-dimensional motion and deformation of the heart wall: estimation with spatial modulation of magnetization: a model-based approach. *Radiology*, 1992, 185, p. 241-247.
109. Ruan, S., A. Bruno, and J.L. Coatrieux, Three dimensional motion and reconstruction of coronary arteries from biplane cineangiography, *Image and Vision Computing*, 1994, 12(10), p. 683-689.

110. Hansis, E., et al., An iterative method for the reconstruction of the coronary arteries from rotational X-ray angiography, Proceedings SPIE Medical Imaging 2007: Physics of Medical Imaging. 2007.
111. C. Blondel, G. Malandain, R. Vaillant, and N. Ayache, Reconstruction of coronary arteries from a single rotational x-ray projection sequence, IEEE Trans. on Med. Imaging, 2006, 25(5), 653–663.
112. Bonnet, S., et al., Dynamic X-ray computed tomography, Proceedings of the IEEE, 2003, 91(10), p. 1574-1587.